

Biweekly Report 2

Comparing dynamic similarity between CRNs using ODE based dynamics

Raunak Narwal

Department of Mathematical Sciences

Indian Institute of Science Education and Research, Mohali, 130406,
Punjab

August 25 to September 7, 2025

Dynamic similarity between Chemical Networks

In the previous report we had already compared static or topological similarities of chemical reaction networks, now we would be exploring the dynamic similarities between them. Dynamics of a reaction can be determined by various ways, some of these ways are listed below.

Deterministic dynamics (ODE based) , each CRN can be expressed as a system of ODEs via mass action kinetics

$$\frac{d\mathbf{X}}{dt} = S \cdot v(\mathbf{X})$$

where \mathbf{X} is the vector of species concentrations, S is the stoichiometric matrix, and $v(\mathbf{X})$ are the reaction rates. Both the networks are simulated under the same initial conditions and the solutions are compared. Time series trajectories are compared using measures like Mean squared error (MSE) between concentration curves, Dynamic Time Warping (DTW) distance (handles phase shifts) and PCA (Principle Component Analysis) on trajectories.

Stochastic dynamics (SSA / Gillespie) , for low copy numbers and noisy networks, both the CRNs are simulated with the Gillespie algorithm (it generates statistically correct trajectories of the system by sampling the time until the next reaction, it determines which reaction occurs based on reaction rates, it works accurately when molecule numbers are low). Distribution of trajectories is compared using Kullback-Leibler (KL) divergence between time dependent probability distributions and Wsserstein distance between ensemble distributions.

Persistent homology of dynamics , instead of building a persistence diagram from the network structure , we can build it from the time evolution data. When we simulate a CRN , we get time series

$$\mathbf{X}(t) = (x_1(t), x_2(t), \dots, x_n(t))$$

This is a trajectory in the state space R^n . Collecting snapshots of this trajectory gives a point cloud (collection of discrete data points in a three-dimensional coordinate system (XYZ), representing the surface of an object or environment). Then Vietories Rips filtration can be used on this point cloud (distance is equal to euclidean between states at different time points) , persistence homology captures the shape of the trajectory in phase space, H_0 : clustering of states (transient vs steady states), H_1 : cylces (oscillations, periodicity) and higher homology tells more complex recurrent patterns. Then compute persistence diargams for both networks, and distances are computed just like we did in biweekly report 1.

In this report, we employed ODE based deterministic dynamics under the assumption of mass action kinetics to compare the time evolution of chemical networks, deterministic ODEs are widely used in systems biology because they give a clear description of how species concentration change in a well mixed environment, without the variability in stochastic fluctuations. Each reaction R_j is represented by a rate law of the form

$$v_j(\mathbf{x}) = k_j \prod_{i=1}^n x_i^{\nu_{ij}},$$

where x_i is the concentration of species i , ν_{ij} is the stoichiometric coefficient of species i in reaction j , and k_j is the reaction rate constant.

$$\frac{d\mathbf{X}}{dt} = S v(\mathbf{X}),$$

where S is the stoichiometric matrix encoding species production and consumption, and $v(\mathbf{x})$ is the vector of reaction fluxes. This deterministic framework allows us to compute the concentration trajectories $\mathbf{x}(t)$ for all species over time, which can then be directly compared between networks. To quantify similarity, we can employ several complementary metrics.

Dynamic Time Warping (DTW) measures the minimal cumulative distance between two sequences after nonlinear temporal alignment:

$$\text{DTW}(X, Y) = \min_{\pi} \sum_{(i,j) \in \pi} \|x_i - y_j\|,$$

where π is a warping path aligning time indices of trajectories X and Y .

Mean Squared Error (MSE) evaluates pointwise deviations,

$$\text{MSE}(X, Y) = \frac{1}{T} \sum_{t=1}^T (x_t - y_t)^2,$$

while Pearson correlation captures linear trend alignment,

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$



Comparing Dynamic similarity between CRNs

while topological analysis give useful insights to use into connectivity patterns, but it fails to capture how a network actually behaves over time. Since chemical and biochemical systems are inherently dynamic, it is critical to analyse how they evolve and whether different networks display similar temporal behaviors. We designed and implemented a python based pipeline that takes two KEGG pathway files in KGML format, then it converts them into dynamical systems under mass action kinetics, simulates their time evolution using ODEs and finally computes and visualizes multiple similarity metrics. The pipeline starts with with KGML parsing, where all the compounds and reactions are extracted. An identifier is assigned to each species and each reaction is parsed into substrate product relationships which are used to construct a stoichiometric representation of the network. Then we generate ODEs under mass action kinetics : for each reaction, the rate of change of species concentration is determined by consumption and production terms in the matrix. Unit rate constants are assumed as there is absense of kinetic data, which allows us to focus on relative behaviors. Then the resulting ODE system is integrated numerically using scipy. matrix of time series trajectories is returned for every species. Then in the similarity analysis stage, we normalize trajectories and compute three complementary metrics, Dynamic Time Warping, Mean Squared Error and Pearson correlation. These metrics are computed for each species individually as well as averaged across all species, ensuring both local and global perspective of similarity. The script generates multiple figures such as the trajectory plots for each network showing concentration time courses of all species, overlay plots where networks are compares directly and heatmaps that depict concentration dynamics as species and time matrices which provides a global landscpae view.

Results for map00400 and map00900

The dynamic similarity analysis of the two networks revealed a mixed pattern of agreement and disagreement in their dynamic behaviors. On average, the Dynamic Time Warping (DTW) distance across the species was high (5.074), which suggests that some species diverged substantially in shape or timing. The MSE (Mean Square Error) averaged 0.33, indicating moderate deviations in concentrations across two networks. The Pearson correlation averaged only 0.186, reflecting weak alignment in overall trends, with several species even displaying strong negative correlation.

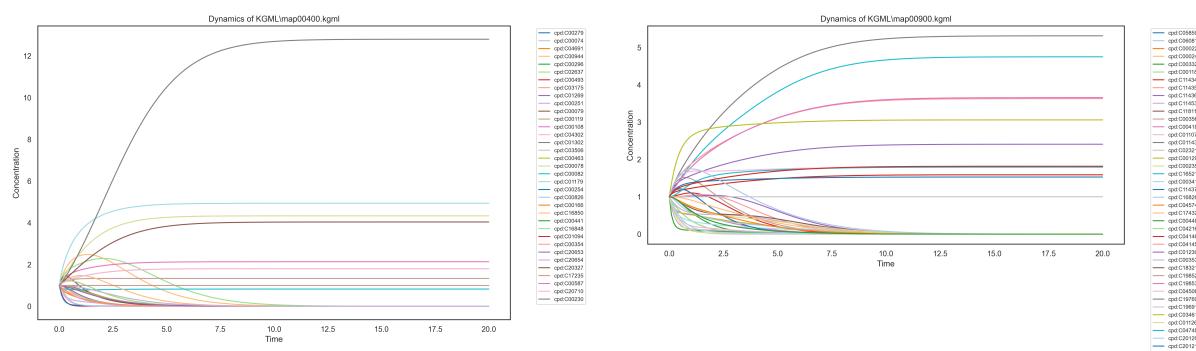
The results were heterogeneous at the per-species level; for example, compounds such

as `cpd:C01302` (DTW = 0.068, MSE = 0.003, ρ = 0.978) and `cpd:C00230` (DTW = 0.214, MSE = 0.004, ρ = 0.997) exhibited near-perfect agreement between the two networks, both in trajectory shape and trend.

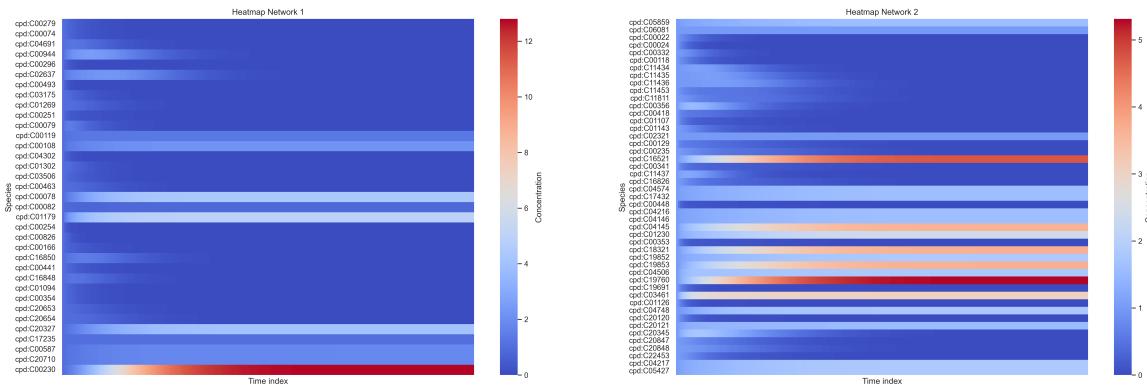
In contrast, species such as `cpd:C00108` and `cpd:C01179` showed very high DTW distances (~ 12) and strongly negative correlations ($\rho \approx -0.9$), which indicates fundamentally different dynamic behavior across networks. This heterogeneity highlights that while certain functional modules or substructures within the pathways evolve similarly, others are dynamically inconsistent.

Output log and Graphs

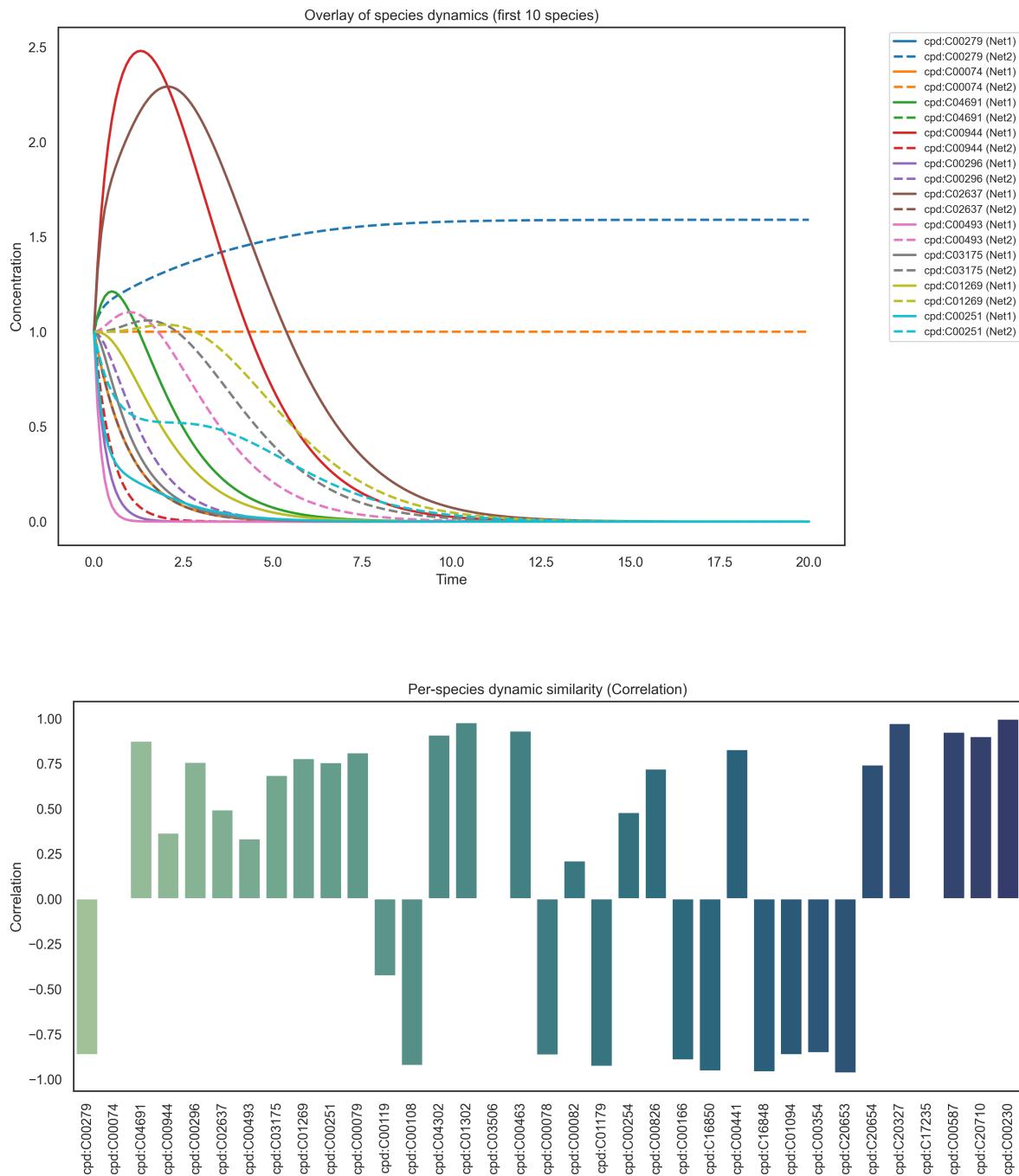
This section contains all the graphs and summary statistics generated by the script. Below chart shows the dynamic change in concentration fo each species in both the networks.

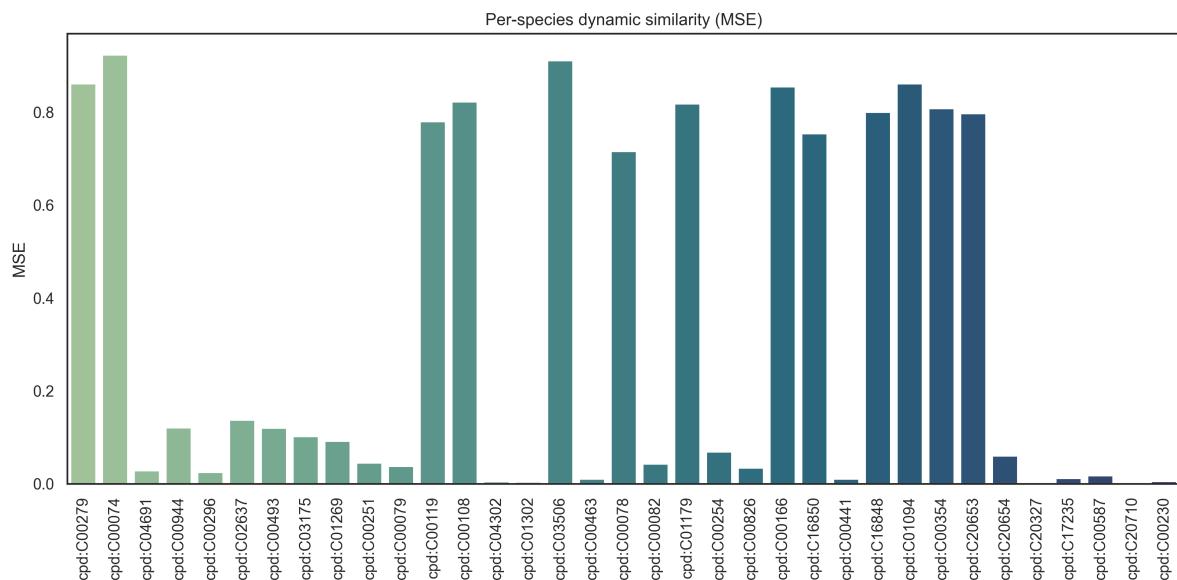
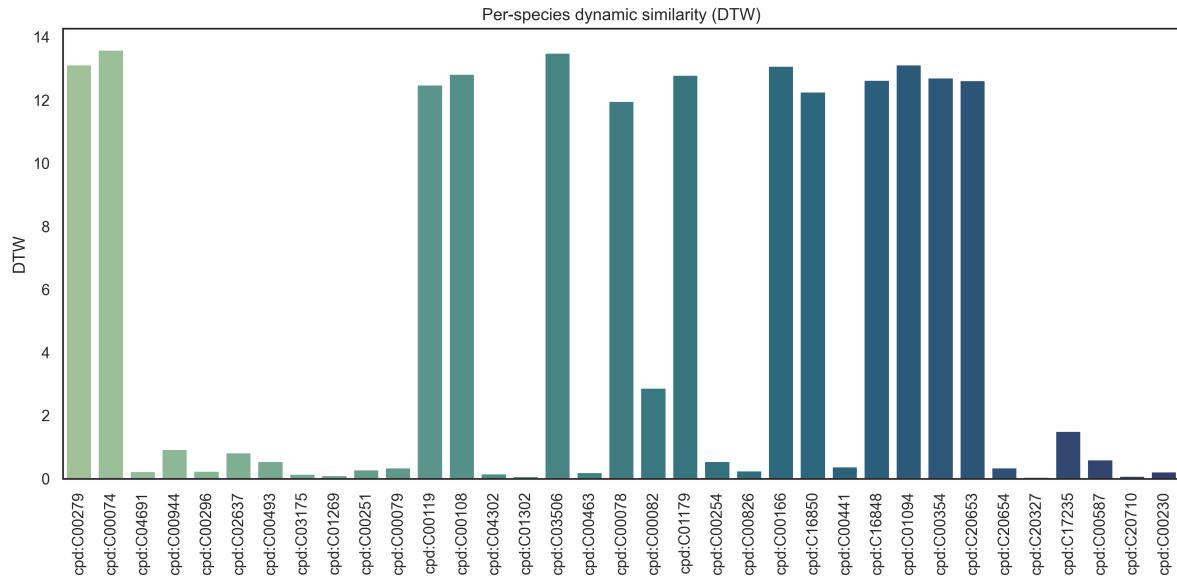


Below is the heatmap version of the line chart.



The overlay chart shows the comparison between the first ten species of each network.





Below is the summary.txt which includes independent DTW, MSE and Corr values for each species.

Dynamic similarity scores:

DTW average: 5.074

MSE average: 0.333

Correlation average: 0.186

Per-species scores:

cpd:C00279: DTW=13.122, MSE=0.861, Corr=-0.864

cpd:C00074: DTW=13.586, MSE=0.923, Corr=nan

cpd:C04691: DTW=0.233, MSE=0.027, Corr=0.875

cpd:C00944: DTW=0.930, MSE=0.120, Corr=0.365

cpd:C00296: DTW=0.240, MSE=0.024, Corr=0.758

cpd:C02637: DTW=0.824, MSE=0.137, Corr=0.493
 cpd:C00493: DTW=0.548, MSE=0.119, Corr=0.333
 cpd:C03175: DTW=0.144, MSE=0.101, Corr=0.685
 cpd:C01269: DTW=0.104, MSE=0.091, Corr=0.778
 cpd:C00251: DTW=0.278, MSE=0.044, Corr=0.755
 cpd:C00079: DTW=0.345, MSE=0.037, Corr=0.810
 cpd:C00119: DTW=12.483, MSE=0.779, Corr=-0.427
 cpd:C00108: DTW=12.822, MSE=0.822, Corr=-0.923
 cpd:C04302: DTW=0.150, MSE=0.004, Corr=0.909
 cpd:C01302: DTW=0.068, MSE=0.003, Corr=0.978
 cpd:C03506: DTW=13.497, MSE=0.911, Corr=nan
 cpd:C00463: DTW=0.197, MSE=0.010, Corr=0.931
 cpd:C00078: DTW=11.959, MSE=0.715, Corr=-0.865
 cpd:C00082: DTW=2.870, MSE=0.042, Corr=0.211
 cpd:C01179: DTW=12.790, MSE=0.818, Corr=-0.928
 cpd:C00254: DTW=0.546, MSE=0.068, Corr=0.480
 cpd:C00826: DTW=0.245, MSE=0.034, Corr=0.721
 cpd:C00166: DTW=13.075, MSE=0.855, Corr=-0.894
 cpd:C16850: DTW=12.263, MSE=0.754, Corr=-0.954
 cpd:C00441: DTW=0.379, MSE=0.010, Corr=0.828
 cpd:C16848: DTW=12.636, MSE=0.800, Corr=-0.959
 cpd:C01094: DTW=13.122, MSE=0.861, Corr=-0.864
 cpd:C00354: DTW=12.711, MSE=0.808, Corr=-0.853
 cpd:C20653: DTW=12.621, MSE=0.796, Corr=-0.966
 cpd:C20654: DTW=0.347, MSE=0.059, Corr=0.742
 cpd:C20327: DTW=0.043, MSE=0.002, Corr=0.973
 cpd:C17235: DTW=1.505, MSE=0.011, Corr=nan
 cpd:C00587: DTW=0.596, MSE=0.017, Corr=0.925
 cpd:C20710: DTW=0.084, MSE=0.001, Corr=0.901
 cpd:C00230: DTW=0.214, MSE=0.004, Corr=0.997

Discussion

Below is the timeline of my work, exams scheduled for 3rd september got cancelled due to floods in Panjab and got rescheduled on 10th September, other deadlines remain the same as before.

Event	Timeline
Biweekly Report 1	21 August
Biweekly Report 1.1	25 August
Mid semester exams	1-2 September
Biweekly Report 2	3 September 7 September
Mid semester exams	3 September 10 September
Biweekly Report 3	21 September

In this discussion section, we wish to discuss what other things we can do with this

persistent homology, dynamic comparisons and discussions upon creating organisms on the local machines. We would discuss upon all the agendas for future work that were mentioned in biweekly report 1 and future agendas of this report as well.

In the previous report, second agenda is not explored yet, it was **comparing permutations of the KGML files using API of KEGG databases** or easier version of this could be , we can save 100s of KGML files in a specific folder, than our script compares these 100 files with each other that would be 4950 comparisons, and return to us just those files that fit our criteria of similarity. We can put a threshold on the error measures and receive new pathways that are similar. This method is faster than manually comparing chemical networks. The API way, lets us not store KGML files on the local machine, rather it calls them directly from the web and runs permutations on the machine. We would surely explore this method in the future biweekly reports.

A **general modular framework** is still lacking, what we need to implement is that we could upload any two kgml files and the scripts would evaluate its topology and dynamics in one run, rather than running and debugging multiple scripts.

Finding substitute/synthetic pathways with similar topology, the idea is to generate synthetic CRNs that preserve the topological signature (persistence diagram) of a target pathway so they can act as functional surrogates or backup pathways.

Node importance / sensitivity analysis using TDA, the idea is to remove nodes and measure how the persistence diagrams change. Nodes that cause large changes in the diagram are topologically critical, those are candidate drug targets or key enzymes. For each node, we can remove it, recompute diagram and compute bottleneck distances. Further we can rank these nodes by change.

Using persistence images as Machine Learning features, we can convert diagrams to fixed sized vectors and use them for clustering , classification and regression, which can be used to predict pathway robustness or drug response.